NASDAQ:AEHR Q2 2026 Earnings Call Transcript

Generated on 8/9/2026

Operator | Conference Operator:

Greetings. Welcome to the Airtel Systems Fiscal 2026 Second Quarter Financial Results Conference Call. At this time, all participants are in listen-only mode. A question and answer session will follow the formal presentation. If anyone should require operator assistance during the conference, please press star zero on your telephone keypad. Please note, this conference is being recorded. I will now turn the conference over to your host, Jim Byers of Fondell Wilkinson Investor Relations.

Jim Byers | Host, Fondell Wilkinson Investor Relations:

You may begin. Thank you, operator. Good afternoon. Welcome to Airtest Systems' second quarter fiscal 2026 financial results conference call. With me on today's call are Airtest Systems President and Chief Executive Officer, Gane Erickson, and Chief Financial Officer, Chris Ciu. Before I turn the call over to Gane and Chris, I'd like to cover a few quick items. This afternoon, right after market closed, Airtest issued a press release announcing its second quarter fiscal 2026 results. The lease is available on the company's website at air.com. This call is being broadcast live over the internet for all interested parties, and the webcast will be archived on the investor relations page of the company's website. I'd like to remind everyone that on today's call, management will be making forward-looking statements that are based on current information and estimates and are subject to a number of risks and uncertainties that could cause actual results to differ materially from those in the forward-looking statements. These factors are discussed in the company's most recent periodic and current reports filed with the SEC. These forward-looking statements, including guidance provided during today's call, are only valid as of this date, and Airtest Systems undertakes no obligation to update the forward-looking statements. Now, with that, I'd like to turn the conference call over to Gane Erickson, President and CEO.

Gane Erickson | President and Chief Executive Officer:

Thanks, Jim. Good afternoon, everyone, and welcome to our second quarter fiscal 26 earnings conference call. I'll begin with an update on the key markets we're targeting for semiconductor test and burn-in, with a particular focus on the common growth drivers we're seeing across these markets, which is namely the massive explosion of AI and data center infrastructure. After that, Chris will walk through our financial performance for the quarter, and then we'll open up the call for questions. While second quarter revenue was softer than anticipated, we made significant progress in both wafer-level burn-in and package-part burn-in segments and are very excited about our prospects moving forward. Based on customer forecasts recently provided to AIR, we believe our bookings in the second half of this fiscal year will be between $60 million and $80 million, which would set the stage for a very strong fiscal 27 that begins on May 30th. During the quarter, we made substantial progress with wafer-level burn-in engagements and production installations across AI processors, flash memory, silicon photonics, gallium nitride, and hard disk drives. We're encouraged to see that one of our key growth strategies focused on reliability solutions for the exploding demand for AI and data center infrastructure is beginning to bear fruit. In package part burn-in, We secured key new device wins for our Sonoma system, supporting high-temperature operating life qualifications for AI devices. These wins are expected to drive additional capacity at test houses, including at least one customer that is elected to move into production in late calendar 26, which we believe could result in meaningful volumes of Sonoma production systems. In addition, in the last month, we received a very large forecast from our lead Sonoma production customer for AI ASIC production capacity. This forecast is expected to drive very strong and

potentially record bookings for the company this fiscal year and position us well for significant revenue growth next fiscal year, with their requested shipments starting in the first fiscal quarter of our next fiscal year. Taken together, our increased visibility across multiple end markets gives us great confidence in our outlook. As a result, we're reinstating financial guidance in fiscal 26, which we'll touch on later in today's call. Now let's talk about our key segments. Starting with our wafer-level burn-in during the quarter, we expanded engagements and completed additional production installations across several end markets. Our lead AI wafer-level burn-in customer continues development of its next generation processor and is currently discussing additional capacity with us. They're forecasting additional system and wafer pack capacity orders this fiscal year and plan to transition to our fully integrated automated wafer pack aligner for 300 millimeter wafers. We expect this customer to continue scaling and excited to support their growth. We also announced a strategic expansion of our partnership with ISE Labs during the quarter to deliver advanced wafer-level test and burn-in services for next-generation high-performance computing and AI applications. This partnership accelerates time to market, improves performance, and gives customers the option of either package part or wafer-level test and burn-in for their production volumes. ISE, together with its parent company, ASE, represents the world's leading outsourced semiconductor assembly and test, or OSAP, platform, serving a global roster of top-tier semiconductor customers. As part of our benchmark evaluation program with a top-tier AI processor supplier we announced last quarter, we completed development of our new fine-pitch wafer packs for wafer-level burn-in of high-current AI processors. These are currently in test with this potential customer's processors and are designed to validate our Fox XP production systems for wafer-level burn-in and functional test of their high-performance, high-power AI processors. We're currently completing startup procedures such as power-up sequencing, thermal profiling, test vectors, timing, and high-speed differential clocks, and expect to complete data collection this quarter. While we're demonstrating our new fine-pitch high-current wafer packs for this benchmark, Many customers can utilize lower-cost wafer pack designs if certain design-for-test rules are incorporated up front. These approaches reduce cost and lead time and are especially attractive to customers focused on faster time-to-market or wafer-level high-temp operating life qualification. We also have two additional AI processor companies planning wafer-level benchmark evaluations since last quarter's earnings call. These benchmarks typically take about six months, and we expect to make meaningful progress beginning this quarter. Both customers are evaluating wafer-level test and burn-in as an alternative to package part or system-level test for large advanced AI modules that combine multiple AI accelerators and stacked high-bandwidth memory. Moving burn-in upstream to the wafer level significantly reduces cost and yield risk by avoiding scrapping expensive substrates and memory stacks when early failures occur later in the process. We have seen estimates that show the cost of the substrate is more than a single processor and the cost of the high bandwidth memory is even higher. Turning to flash memory, we completed our wafer level benchmark with a global leader in NAND flash just prior to the holidays. The customer has now taken the wafers back for further processing to validate correlation with their internal process. This benchmark demonstrated our ability to test flash memory wafers with significantly higher parallelism and power than is possible using traditional probers and group probers from companies such as TEL or Acratech. We've also proposed the next generation solution enabling test of a new emerging flash memory device called High Bandwidth Flash or HBF designed for AI workloads. This proposed solution leverages our Fox XP platform, wafer packs, and auto-aligner technology, and would support single-touchdown, high-power tests on 300-millimeter wafers. While development of this system would take over a year following customer commitment, we believe this represents a compelling entry point into a large and evolving memory market. We look forward to sharing more details as this progresses. Turning to silicon photonics, we believe that silicon photonics used in data center and also chip-to-chip I.O. is going to be a significant market driving production burning capacity for our Fox wafer-level burning systems and wafer packs. Our lead customer has now firmed up its production ramp, which we expect to begin early next fiscal year. While this timing is later than previously expected, it aligns with recently announced AI processor platforms and positions as well for calendar 2026 orders and deliveries in fiscal 27. We've also finalized a forecast with another major Silicon Photonics customer, initially targeting data center applications with a roadmap toward optical IO. We expect to book their initial turnkey FOX system soon with delivery planned for May of this year. In gallium nitride power semiconductors, We continue to support our lead production customer, though we experienced delays related to unanticipated high-voltage bolt conditions that required

wafer packs and protection circuit redesigns. This delayed approximately $2 million in wafer pack shipments from last quarter into this quarter, along with some system enhancements. Shipments have now resumed, and lessons learned have significantly strengthened our GAN power supply burning capabilities. If anyone tells you that testing and burning in full wafers of GaN power semiconductors with up to 600 volts or more is easy, don't listen to them. We also continue to engage with multiple new potential GaN customers and are developing wafer packs for several new device designs that are expected to go to high-volume production for applications like data center infrastructure and power delivery automotive electrical power distribution on both ICE and hybrid electric vehicles, and even power semiconductors used for electrical breakers. AIR has a unique solution that can deliver full turnkey, fully automated wafer handling and probing for test and burn-in of GaN wafers in sizes from 6 to 8 inches and even 12 inches or 300 millimeter wafers. Turning to silicon carbide, as we previously discussed, silicon carbide demand has been weighed toward the end of this fiscal year. Customers continue to be optimistic about this market and their capacity needs. But we've tried to take a very conservative stance that is mostly show us the orders before we believe them. Our lead customer recently transitioned from 150 millimeters to 200 millimeter wafers, nearly doubling output without adding new Fox XP systems and supported by AERA's proprietary wafer packs that we developed to accommodate both 150 and 200 millimeter wafers, contacting 100% of the dye on each in a single touchdown. They're now seeing additional needs for wafer packs this year, but additional capacity for systems appears to be a year out. We pushed out expected orders until next fiscal year from our near-term forecast, but have capacity of systems or wafer packs to continue to support their surge capacity needs as well as our other silicon carbide customers. While electric vehicle-related demand has slowed industry-wide, we remain well-positioned with the most competitive wafer-level burn-in solution available, and we expect to benefit when growth resumes. In semiconductors used in data center hard disk drives, we're installing the additional Fox CP systems for a major supplier of hard disk drives for wafer-level burn-in of their special components in their drives. They've indicated plans for additional purchases later this calendar year. While their device unit volumes are very large, the overall revenue opportunity remains modest due to short stress times and the massive parallelism achieved on our Fox CP system and proprietary high-power wafer-packed wafer contactors. Now, let me talk about package part burn-in. We're seeing continued momentum in package part qualification and production burn-in for AI processors, driving growth in our new Sonoma ultra-high power package part burn-in systems and consumables. As we announced today in a separate press release, during our fiscal third quarter to date, we have received orders from multiple customers, totaling more than $5.5 million for our Sonoma ultra-high power package part burn-in systems, including initial orders from a premier Silicon Valley test lab for our newly introduced higher power configured Sonoma system that can also support full automation. These orders already exceed the total Sonoma orders for the entire second quarter, highlighting the accelerating demand we're seeing for our package-level burden of high-powered AI and compute devices. This quarter, we also secured key new device wins on the Sonoma platform, or high temp operating life qualification. These wins are expected to drive additional capacity at test houses, with at least one customer planning to transition to production later this calendar year, generating significant system demand. Our lead package part burn-in production customer for AI processors continues to ramp and is forecasting substantial growth in 2026 and beyond. Although we have not yet received the purchase order, we have received a substantial forecast from this customer for AI ASIC production capacity, with requested Sonoma production, package part burn-in system, and BIM shipments beginning in the fiscal first quarter of 27. That starts May 30th, which we expect to contribute to very strong bookings in fiscal 26 and generate significant revenue growth in fiscal 27. This customer also plans to introduce much higher power A6 later this year, for which we are already developing the high-temp operating life qualification burden modules and sockets to be used on the Sonoma systems at one of the premier Silicon Valley test services companies that have many systems installed. This AI accelerator ASIC processor is also forecasted to go to production burden and drive even higher volume needs for production burden systems downstream at the OSATs in Asia. We feel we're very well positioned with our Sonoma system for this production capacity need and believe this could drive very substantial volumes of Sonoma systems in our next fiscal year. During the quarter, we completed development of a next generation fully automated higher power Sonoma system supporting up to 2,000 watts per device. This system enables continuous flow operation, improved throughput, and seamless transition from qualification to high volume production using the same fixtures and sockets. These capabilities enable customers who are

focused on high-temp operating life reliability testing to have a system that is fully software and hardware compatible with the Sonoma systems they have installed, which simplifies and accelerates time to market that's critical for HTAL testing and new AI processors. This Sonoma burn-in system can also simply bolt on a fully automated handler developed and sold by Airtest as a turnkey solution to allow hands-free operation with less than a couple of minutes of overhead per burn-in cycle, which is amazing for production burn-in needs. We're also seeing increased demand for our lower-power Echo and Tahoe package part burn-in systems, driven by our installed base of more than 100 systems across over 20 semiconductor companies worldwide. But I'll wait for another call to discuss these systems and the markets they serve in more detail. As stated last quarter, the rapid advancement of generative AI and the accelerating electrification of transportation and global infrastructure represent two of the most significant macro trends impacting the semiconductor industry today. These transformative forces are driving enormous growth in semiconductor demand, while fundamentally increasing the performance, reliability, safety, and security requirements of the devices used across computing and data infrastructure, telecommunications networks, hard disk drive and solid state storage solutions, electric vehicles, charging systems, and renewable energy generation. As these applications operate at ever higher power levels and in increasingly mission-critical environments, The need for comprehensive test and burn-in has become more essential than ever. Semiconductor manufacturers are turning to advanced wafer-level and package-level burn-in systems to screen for early life failures, validate long-term reliability, and ensure consistent performance under extreme electrical and thermal stress conditions. This growing emphasis on reliability testing reflects a fundamental shift in the industry. from simply achieving functionality to guaranteeing dependable operation throughout a product's lifetime, a requirement that continues to expand alongside the scale and complexity of next-generation semiconductor devices. This year, we're making significant progress expanding into additional key markets for our semiconductor test and burning solutions, including AI processors, gallium nitride power semiconductors, data storage devices, silicon photonics, integrated circuits, and flash memory. This diversification of our markets and customers is significant, given our revenue concentration in silicon carbide for electric vehicles the last two years. This progress and key initiatives expands our total addressable market, diversifies our customer base, and provides us with new products, capabilities, and capacity, all aimed at driving revenue growth and increasing profitability. The progress we made this quarter with a significant number of customer engagements and production installations provides improved visibility into future demand. As a result, we're reinstating guidance for the second half of fiscal 26. For the second half of fiscal 26, which began November 29th, 25, and ends this May 29th of 26, AIR expects revenue between $25 million and $30 million. As stated earlier, although we're not providing formal bookings guidance, based on customer forecasts recently provided to AIR, we believe our bookings in the second half of this fiscal year will be much higher than revenue, between $60 million and $80 million in bookings, which would set the stage for a very strong fiscal 27 that begins on May 30th of 2026. With that, let me turn it over to Chris, and then we'll open up the lines for questions.

Chris Ciu | Chief Financial Officer:

Thank you, Gane, and good afternoon, everyone. I'll begin with bookings and backlog, then walk through our second quarter financial performance, cash position outlook, and investor activity. The company recognized bookings of 6.2 million in the second quarter of fiscal 2026, compared to 11.4 million in the first quarter. At the end of the quarter, our backlog was 11.8 million. Importantly, during the first six weeks of the third quarter, we received an additional 6.5 million in bookings. This increase was driven primarily by an order from a premier Silicon Valley test lab for our newly introduced high power configured Sonoma system, which we announced this afternoon. Including this recent bookings, our effective backlog has now grown to 18.3 million, providing increased visibility as we move through the remainder of fiscal 2026. Turning to our second quarter results, revenue was 9.9 million, down 27% from 13.5 million in prior year period. The decline was primarily driven by lower shipments of wafer packs, partially offset by stronger demand for our Sonoma systems from our hyperscaler customer. Contacted revenues, which include wafer packs for our wafer-level burn-in business and BIMs and BIPs for our packaged part burn-in business, totaled 3.4 million. representing 35% of total revenue. This compares to 8.6 million or 64% of revenue in the second quarter last year. Non-GAAP growth

margin for the second quarter was 29.8% compared with 45.3% a year ago. The year-over-year decline reflects lower overall sales volume and a less favorable product mix as last year's quarter included a higher proportion of higher margin wafer pack revenue. Non-GAAP operating expenses in the second quarter were 5.7 million, down 4% from 5.9 million in Q2 last year. The decrease was primarily due to lower personnel related expenses, which were partially offset by high research and development costs, including higher project spending, as we continue to invest resources in AI benchmark initiatives and memory related programs. As previously announced, We successfully closed the in-cow facility on May 30th, 2025, and completed the consolidation of personnel and manufacturing into its Fremont facility at the end of fiscal 2025. During the quarter, we negotiated an early lease termination with the landlord, reducing our obligation by five months of rent. As a result, we recorded a reversal of $213,000 related to a previously accrued one-time restructuring charge. During the quarter, we recorded an income tax benefit of 1.2 million, resulting in an effective tax rate of 27.3%. Non-GAAP net loss for the quarter, which excludes the impact of stock-based compensation, acquisition-related adjustments, and restructuring charges, was 1.3 million, or negative four cents per diluted share, compared to net income of 0.7 million, or two cents per diluted share in the second quarter of fiscal 2025. Turning to cash flow, we used 1.2 million in operating cash during the second quarter. We ended the quarter with 31 million in cash, cash equivalents, and restricted cash, up from 24.7 million at the end of Q1. The increase was primarily due to proceeds from our at-the-market equity program. As a reminder, in the second quarter of fiscal 2025, we filed a new 100 million S3 self-registration that was approved by the SEC for three years. followed by an ATM offering of up to $40 million. During the second quarter fiscal 2026, we raised $10 million in gross proceeds through the sale of about 384,000 shares. At quarter end, 30 million remained available under the ATM. We intend to utilize the ATM selectively with a disciplined approach focused on market conditions and shareholder value. Looking ahead to the second half of fiscal 2026, which began on November 29th, 2025 and ends on May 29th, 2026, we expect total revenue between 25 million to 30 million and non-GAAP net loss per diluted share between negative nine cents and negative five cents for the six month period. On the investor relations front, last month on December 17th, 2025, Lake Street Capital initiated annual research coverage on Airtest, along with equity research firm Freedom Broker, which initiated coverage last June. There are now a total of four research firms covering the company. Lastly, look at the investor relations calendar. We will meet with investors at the 20th Annual Needham Growth Conference in New York on Tuesday, January 13th, and then return to New York in February for the 15th annual Susquehanna Technology Conference on Thursday, February 26th. We'll also be participating virtually in the Oppenheimer Emerging Growth Conference on Tuesday, February 3rd. We hope to see you at these conferences. That concludes our prepared remarks. We're now happy to take your questions. Operator, please go ahead.

Operator | Conference Operator:

Thank you. At this time, we will be conducting a question and answer session. If you would like to ask a question, please press star 1 on your telephone keypad. A confirmation tone will indicate your line is in the question queue. You may press star 2 if you would like to remove your question from the queue. For participants using speaker equipment, it may be necessary to pick up your handset before pressing the star keys. One moment, please, while we poll for questions. Once again, please press star 1 if you have a question or comment. Our first question comes from Christian Schwab with Craig Hallam. Please proceed.

Christian Schwab | Analyst, Craig-Hallum Capital Group:

Again, thanks for all the details on the call. What wasn't clear to me exactly is on the booking strength, potential booking strength of 60 to 80 million in the second half of this fiscal year. Is that almost entirely on the AI accelerator processor line?

Gane Erickson | President and Chief Executive Officer:

There's some silicon carbide, not much, like not very much at all. There is some silicon photonics for sure, but the bulk of it is across wafer level and package part burn-in for AI processors, yes.

Christian Schwab | Analyst, Craig-Hallum Capital Group:

Okay, perfect. And then, given that such a material bookings from the AI processor market, can you give us any indication or idea, you know, I know we've talked about the opportunity of that marketplace being bigger than silicon carbide, but, you know, what's narrow it down to, you know, kind of a multi-year time frame, kind of including 27 and 28. Do you see that business after initial orders expanding meaningfully from there?

Gane Erickson | President and Chief Executive Officer:

We do. We do. We do. And we've been taking a pretty conservative stance – On how large, particularly AI and the wafer-level side of it is, and conservative may not be fair, candidly, we're still trying to get our arms around how big it is. What we get is visibility of a specific, you know, GPU or CPU or, you know, network processor or an ASIC. And then You know, we hear these things from the customer, and then we look externally, and what are they telling the street, and try and correlate through those lookups. And I'd say pretty consistently we hear bigger numbers from the customer than the street. Not sure what that all means, okay? And then as they give us test time estimates of what the burning conditions are, we can start to put some numbers around it. But, you know, a single processor – So for some of these big guys at wafer-level burn-in is, you know, 20, 30 systems or so. And these are $4 million or $5 million machines. So you get a feel for the size of what that looks like. And, you know, the, you know, estimates of, you know, today, if you were to look at AI spend in test between test and burn-in, you know, is it $8, $10 billion? Yes. maybe $15 billion or so. I mean, it's a really large number. So we want to get ahead of ourselves here. But, you know, when customers ask you things like, how many can you make? So can the AI business be measured in hundreds of millions of dollars for air test, you know, a few years out? Yes, for sure. Now, what's interesting is that we're in this – I think it's an awesome position to be in because the Sonoma system, you know, is a highly preferred system for HTAL, the high-temp operating life reliability testing for these AI processors. It has the largest installed base in all the test houses around the world. We're getting people that approach us because we are like, I don't want to say we're the de facto standard, that's probably bold, but we have more capacity than everybody else, and therefore they are saying you're kind of the go-to guy. I like those words. And we can build lots of them. So customers are using that, and we get a front row seat to actually bring them up. Then we say, oh, by the way, you know, if you want, you can take this machine, add production handling to it, and do production on it. In the meantime, if you come to our facility and you do a tour and you can see that production test cell for the Sonoma automation, we, of course, will walk you by a Fox wafer-level burn-in test cell and mention, oh, by the way, that happens to be doing a benchmark on a 300-millimeter wafer. We can't tell you who it is. And so they're like, well, what is that? So we're in a position to be able to talk about both of them. and the ASPs are actually higher on the wafer level side of things, but the value proposition way outweighs that because of the yield advantage of doing it wafer level. The yield savings dwarfs any of the costs of the cost to test the wafer level burden. So as we get our arms around the market, the market data that would be out there would be package part because no one's doing wafer level except for us. And so we're creating our own models related to, okay, for that unit capacity, if you went to wafer-level burn-in, what would that look like? Kind of similar to what we had to go through in the original silicon carbide side of things, you know, if the whole market. And we're not saying, you know, everybody included, NVIDIA and Google and Microsoft and Tesla and these guys all went with us. How big is that market? We haven't even tried to put our arms around that yet. But it's substantial.

Christian Schwab | Analyst, Craig-Hallum Capital Group:

Great. And then I guess one last question, if I may, and follow up on your comment about capacity. You know, how many systems do you think you're capable of manufacturing in a year for wafer level?

Gane Erickson | President and Chief Executive Officer:

We have talked to customers about capacities exceeding 20 systems a month. at either package or wafer level. If we had to, we could ship 20 systems a month of each during this calendar year. Now, that's bigger than our forecast by a lot. But you know what? When people are saying, could you do something like this and intercept something, it's like if they gave you an order for 50 or 100 Sonomas, how long is it going to take you to build them? Makes sense.

Christian Schwab | Analyst, Craig-Hallum Capital Group:

Makes perfect sense. No other questions. Thanks, Gabe. You're welcome.

Operator | Conference Operator:

The next question comes from Jed Dorsheimer with William Blair. Please proceed.

Unidentified Participant:

Hey, Jed. Oh, Jed.

Gane Erickson | President and Chief Executive Officer:

We got you on mute, Jed.

Jed Dorsheimer | Analyst, William Blair:

Oh, okay. I was that guy. So, anyways, you got me. Oh, there you go. Yep, yep. All right. So thanks for taking my question. Yeah, I guess maybe just to start, on the wafer level, you know, I think your prior comments around the timing of the benchmark, it seems like that's taken a little bit longer. And I'm just wondering, you know, is that a function of – is it because it's new and what you're seeing is from the customer is that they're changing parameters or that's extending that out? Because I think you had maybe talked about, you know, by February timeframe, and we're almost, you know, we're... Do you want me to throw my customer under the bus?

Gane Erickson | President and Chief Executive Officer:

Is that what you're trying to tell me? No, no, no. No, no, no, no, no. Let me answer that. No, I got it. I got it. No, it's totally fair, okay? What I do in all of these things is try to describe exactly what we feel, what we know, what we knew at the time. One of the things that's very interesting and fun about this particular customer who is a very notable customer. When they gave us, and I don't think I've overstimulated, when they gave us the vectors, the test vectors, etc., they were giving it off of a platform from package level. Package and wafer are different. We had a huge arm wrestle with them related to what they could actually do at wafer level and also to demonstrate to them significant DFT, lower pin count modes, et cetera, to be able to

do it at wafer level, which was a big deal because they never understood that because, of course, nobody's ever done this before with us. I'll just leave it at this. They actually gave us some things that were implied based upon package that weren't totally applicable to wafer level, and we struggled with some of that. And it turns out, so it actually did delay a little bit. I think it's mutually understood. It's like, oh, sorry, you know. We were thinking in package, we forget about wafer and sort. And that's a growing thing. We've seen this with other customers. On the very first time you're doing wafer-level burn-in, you just don't think about it from... the challenges or the differences at what happens when you're talking about a device that shares common substrates or, you know, from a probing environment. So is it longer? Maybe a little bit, you know, measured in, you know, weeks or a couple months or something. But, you know, some of the things that, like, mechanically, the way for physical contact to the device, to using our auto aligner to pack these new, fine pitch wafer packs, the test plan itself, the vectors, those things were all going along pretty well. So I wish it was a little bit sooner, but I think we're still very much on track to try and get them some data over the next, you know, couple months here, or even maybe even this month. So now the question, of course, parlays into what do they do with it? What's the timing? You know, do you understand what device they want to cut in? We do. We're not going to share that with you guys. You know, are we going to make it? We believe we're still, you know, there's lots of reasons to actually want to cut in wafer-level burn-in. And the sooner, the better. So I'm actually, you know, we're really excited about this particular one. And then now we've got another couple guys that are saying, pick me, pick me, too. and are generating the information to give us so that we can actually do design reviews and walk through a wafer pack design for them as well.

Jed Dorsheimer | Analyst, William Blair:

Got it. That's helpful. Thanks. And I just want to address the potential of cannibalization between package and wafer level. And if I read through your comments, it seems like the – AI processor is what's moving along with this customer on the wafer level. You had mentioned briefly, actually, on the ASIC side. Are you anticipating that the ASICs basically run with package level and that AI processors are wafer level, or are you anticipating both at wafer level?

Gane Erickson | President and Chief Executive Officer:

Thanks. Thanks. Yeah, okay. So vocabulary for everybody that's listening out there, right? So when you talk about processors and the AI, you know, arguably there's even maybe at least two or three different broad flavors of them, okay? You're going to have the actual GPU if it's an NVIDIA or ASIC when you talk about everybody else's. In reality, the GPU is kind of an ASIC at NVIDIA too. Jensen said that at one point. These are AI accelerator platforms, okay? And they can be used for language models or for inference type things. There's also processors that like CPUs, like Intel or Grace or Vera type CPUs and others that are making them. that are also going through a burn-in process. And then you could argue there's even network processors and things like that. But generally, when we talk about AI processors, we're generally in the CPU and GPU type or ASIC type that are combined together in these AI processor, you know, clusters. And things like you hear at GB200 is Grace CPU and two Blackwell processors AI accelerators in one package, if you will, or in one cluster. What's happening with the roadmap is that devices are going from, you know, a single AI accelerator or CPU in a package to a package that includes embedded memory, like high bandwidth memory and high bandwidth flash over time. and then to having more than one compute chip in it, so having two processors in it or four or eight, like you look at the Intel or the AMD roadmap. Everyone has a roadmap to two or four more AI processors on a single substrate. What's happening is that there is a – the qualification of those are all done today in a full package. The whole device in a big substrate is done, and it can take months to even go to get the packaging to qualify that. So there are people that would like to be able to qualify the processor inside when it's still in wafer form, right? From a production perspective, the value proposition is you're burning in these devices – And when they fail, you take out the other compute chip and all the memory plus the co-auth substrate, which costs more than the silicon of the compute chip itself. So the roadmap is getting more intense. So there's people that are like, oh, I want to evaluate this for this device.

This would make sense. But, boy, that next one makes twice as much sense, and the one next to that is four times as much sense because of this evolution. So, you know, a lot of times we discuss, okay, is there a window? Like what happens if you just miss this one device? It doesn't feel like that. It's a treadmill of you can always step on. And the customers are like, okay, how do I cut you in? I've said publicly that our large package part production customer, we've talked about it as an ASIC hyperscaler, they're actually on Sonoma production. We're qualifying their next device that's going to go to production, we believe, and hope it'll go on Sonoma as well. The third one they're giving us design files of so we can make sure that Sonoma is ready for that. But they've also said, you know what, by then maybe we'll want to consider FoxWave's a little burn-in. And the interesting thing is, it's like, well, what will you do with all the package systems from us? You know, who cares? You know, it's like, what? Because if I can move it to wafer level... I don't need to do it in package anymore. Now, will it cut over just like that? We'll see. I think the world's going to be both for a long time, and we're in a great position to do both. But is there cannibalization? For sure. We had a customer come in who wanted to talk about what we thought was packaged for burden. Alberto, our VP over the package for our business, and I met with them. And 15 minutes into the meeting, he goes, I'd like to talk about wafer level. And Alberto looked over at me, and I'm like, okay, new slides. You know, it's like, so at least we got both. And, you know, we're in a great position. And actually, I would say all three. We do the high-temp operating life today only at package over time at wafer level. And we do production burn at either package or wafer level. So, a great front row seat.

Jed Dorsheimer | Analyst, William Blair:

That's helpful. I'll jump back in the queue. Thanks.

Gane Erickson | President and Chief Executive Officer:

Okay. Thanks, Jed.

Operator | Conference Operator:

Our next question comes from Max Michaelis with Lake Street Capital. Please proceed.

Max Michaelis | Analyst, Lake Street Capital:

Hey, Max. Hey, guys. Thanks for taking my question. First one for me, just around the bookings guide, I know you previously shared that's majority of around AI. But just given the distinction between the low end and the high end, if we just take the midpoint around 70 million, I mean, to get to that 80 million, is that all basically around AI or does that suggest any improvement around silicon carbide or GaN?

Gane Erickson | President and Chief Executive Officer:

You know, it's the least in that number is silicon carbide, okay? And then GaN is pretty close. Hard disk drive's a little bigger than silicon photonics is a chunk. I mean, we've got production systems in there for our lead customer. We have a new customer that wants a system. They want it shipped by May. We're, you know, suggesting to them that they really should get their order in before we ship it. Joke, joke. I'm kidding. It's a challenge right now because they're like, please, please build it. We actually have a system on our floor, and if they get their PO in, if you're listening, you get to get it. If not, we'll give it to the next guy. But anyhow, and then it would be wafer-level burn-in, and then I think package is the biggest. I'm sorry, wafer-level burn-in of AI, and then package part AI is the biggest.

Max Michaelis | Analyst, Lake Street Capital:

Okay. So, yeah, that just, I suggest the $16 million, $80 million suggests just greater volume. orders from wafer-level burn-in. Okay. And then lastly, I haven't had time to run through the entire press release, but that $5.5 million order you noted in your prepared remarks, can you share some more detail on that? Is there anything new that we should be looking for that's just kind of standard?

Gane Erickson | President and Chief Executive Officer:

You know what? It has a mix of some customers that already had Sonomas that were buying more. that were AI-related. It had some far-end modules that was important because it was for a new design of a really expected-to-be high runner that's going to production. It has a big order from what we call a premier Silicon Valley test services company. We'll leave it at that. They actually bought a number of the new Sonoma configurations which are the very high power ones that allow them to go to 2,000 watts. We have some devices that we're going to be testing this spring that are almost 2,000 watts per device, right? And everybody's out there talking about how can you do, you know, what does it take to get to 1,000 watts? We're jumping right past that. And this is in a high-volume Sonoma system, so they'll be able to test a large number of devices in that system And I think the numbers – I should know this number. I think it's 44 devices. But, I mean, it's a large number of devices to be able to test those. And it's either – by the way, it's either 22 or 44. I should know that. Sorry, folks, to go through the math on that particular application because of the number of resources and power supplies and things. But it's the biggest part we've seen that's in development, and that's going to be going to production. So that's a big deal. So it's a combination of several different orders. Every one of them is kind of sort of strategic to us.

Max Michaelis | Analyst, Lake Street Capital:

All right. Thanks for taking my questions. You're welcome. Thanks, Max.

Operator | Conference Operator:

The next question comes from Larry Shlebina with Shlebina Capital. Please proceed.

Larry Shlebina | Analyst, Shlebina Capital:

Hey, Larry. We try to line up your ramp or at least your demand for the systems that you're working on developing for these customers on the AI processors with what's publicly disclosed in terms of the product launch. Is there a case where they may start up on package part, wherever they have the capacity to do that, and then when they feel comfortable, maybe if it's after the product's launched, Would they cut over the wafer-level burn-in because it's so much more efficient and saves them money? Would they do that, or would they just do it initially on a brand-new product launch at the beginning?

Gane Erickson | President and Chief Executive Officer:

That's kind of – do you have a sense of that? Okay, so there's two things in there. What I definitely see happening is, you know, we know for a fact a customer was doing system-level – a rack test, okay – And the only time they identified infant mortality or early life failures was when it's installed in the data center. Pretty nasty, okay? That's test or not or burn. And so they said, we'll run it for two weeks, and if it hasn't died, we'll accept it, you know, kind of thing. And then they'll actually plug it into the network. Pretty expensive way of doing it. Then there are companies like AEM and Advantest and Teradyne that have talked about

System-level test machines, which is a type of ATE machine that is designed to be doing a high-speed insertion and boot up, like, the operating system, it's a great way to do a very high degree of test coverage for a specific application. People were saying, oh, we're going to do burn-in with that. Well, that doesn't really, you know, those systems are designed for high speed. They're designed to be at the user mode. They're designed to run cold. They're not really designed for burn-in, and they're quite expensive and large. But the market was pulling on that because it's sure better than doing it in Iraq. And there wasn't another system available in what a lot of people refer to as ovens, which is a large-scale system that you put lots of burn-in modules or trays with lots of devices and test all at once. Those were like from KYC or something, maybe 600 watts and below or something, and there really wasn't a tool out there for that. This is where Sonoma was pulled up because we were doing, NCAL was using it for the high-temp operating life, but it's like, well, wait a minute, can I use that in production? Can you add automation? Can you do these things, support, and can you, you know, quadruple or, you know, 50x your capacity? So that's where Sonoma is coming in. When Sonoma enters that market, doing system-level test or rack test makes no sense whatsoever. So it's highly competitive as that. Now, having said that, wafer-level burn-in is even better. But a lot of people may say, well, I need to think through that. You know, where do I put that insertion? You know, I might need to implement some design for test modes to be able to implement it, at least to take advantage of the very low-cost, you know, full wafer contactors from Airtest and things like that. So I think it's an evolution, but I think – You know, the conversation we have with customers is they're like, I need package-bar burn-in. Let's talk about that. But, boy, wafer-level burn-in would be better. How do we engage on that? And then specifically on a per-customer basis, you know, I don't want to get too carried away with our strategy, but if you have an installed base of something, you know, package-bar burn-in systems or, you know, I could go in and displace you with maybe Sonoma and But it's probably better for me to go displace you with wafer-level burn-in because it's not even a price thing in that sense. It's yield or capacity. So it depends on the customer, and we have some customers that have some devices that want to think about wafer-level, some they want to think about package, some they want to think about package, and then eventually the wafer-level over time. Okay. I hope that wasn't – as I look back, that was pretty confusing. But, you know, there's – it's an evolution of it, and, you know, guess what we do? The customer's always right. You tell me what you want, and, you know, we're in.

Larry Shlebina | Analyst, Shlebina Capital:

Well, if you're – all these evaluations they have going on with wafer-level burning, if it takes longer and the product ends up getting launched, would they still cut over to some portion of the production – on wafer-level burn-in, once it's proven out for the particular product or the predicted, would they do that midstream?

Gane Erickson | President and Chief Executive Officer:

I think it depends. It's not a slam dunk. I mean, I think traditionally people will start a product and, you know, do the release of that one product on one test platform or something, and then you cut in on the next one. I think that'd be fair to say that. But there are certain devices we know that their intended application, there's two or three different applications for it. So, you know, for a lot of language model, maybe they think about it one way, but if it's going to be automotive, then that's a different thing, right? So even within a product, there might be an evolution. Or they get by until they can implement wafer-level burden. That particularly comes in the fact when you think about a multi-chip module. As soon as you could do wafer-level burn-in, if I could save you 1% yield per die on a four-die AI processor that has a $15,000 bomb, of course you would do that, right? I'm not sure if they would. We're trying to be as open as we can. We know as much as we know, but... You know, there's definitely advantages to do wafer level. I mean, ultimately, that's the most, you know, kind of the best place you could ever do it. And if you implement some DFT and you implement some of the things we do, I could build you a wafer pack in eight weeks. Have you on wafer?

Larry Shlebina | Analyst, Shlebina Capital:

The shift gears on the flash benchmark that you completed right now a little bit ago before the holidays, what do you expect the customer to get back to you and, you know, And more importantly, when do you expect them to come with an order?

Gane Erickson | President and Chief Executive Officer:

I was waiting for somebody. Yeah, that's where my head's at, too. My guess is, Larry, the next couple months or so for them really to get back, depending on how they – the wafer's going back to test, which is tested at wafer. I don't think they're going to package it up and go through some stress qualification test. That might be something. But, you know, we've already had some design reviews with them on our new tester and planted the seeds. They were very impressed is how I would describe it. The big shift here was, you know, when we even started this thinking to do the benchmark with them, which is, what, like a year ago. Kai, if I get that right. No, a year and a half ago. Yeah, yeah, fair enough, right? When we were starting to even build up to get the design files and what wafer we were going to be testing with them, it was not aimed at high bandwidth flash because that didn't even exist, right? They were looking at it for, like, commodity, you know, data center SSDs. Now with the HBF, it broke their infrastructure, the power supplies, IO pins, et cetera, and parallelism, And now they have a power problem, which we love. Well, we're good at power. So people that have power problems, that's music to our ears. So, yeah.

Larry Shlebina | Analyst, Shlebina Capital:

If I recall, you originally said the driver, their motivation was as the 3D NAMs got higher levels of, you know, they're even talking about getting to 400 levels.

Jim Byers | Host, Fondell Wilkinson Investor Relations:

Layers, layers, yeah.

Larry Shlebina | Analyst, Shlebina Capital:

Layers, yeah. That required more power and exceeded the power in their existing system. That's right. So that they need your high power. So here we are a year and a half later, and so how are they getting by to this point? And don't they need your high power capability?

Gane Erickson | President and Chief Executive Officer:

They're having to – they can't test the whole way for one touchdown, as an example. Mm-hmm. But what I described there, which people, if you follow along with that, that was actually referred to as hybrid bonded flash. Same letters, by the way. Hybrid bonded flash was a novel idea that the base substrate layer was logic, done on a logic process, and then you build up just the stack memory, and you do that in a memory process, and then you bond them together. The result of it is that memory stack is a taller building with a smaller footprint, so you get more die per wafer. That's good, right? But the power was much higher. HPF, as in high bandwidth flash, is in some ways architecturally similar, except for it's more power. Because of its speed, it has additional power supplies. and it's taller, it actually is even more of a problem for them, which, you know, I guess if you're a tester guy, the bigger the problem, you know, you have more to solve. But we had to go back and redesign the tester because we were originally aiming it at the other device.

Larry Shlebina | Analyst, Shlebina Capital:

I would think they would need more capacity for the enterprise flash part of it before they ever start testing. needing something for HBO. So the Enterprise Flash, I'm wondering, you know, when is something going to happen there? It seems like it's overdue.

Gane Erickson | President and Chief Executive Officer:

Yeah, I mean, our goal in this case would be, you know, we had originally hoped to finish the benchmark at the end of last year, okay, so like, you know, we're six months later, and I think as I shared with you, if you read through all of the notes, around March, it was like it felt like you're pushing a rope. Something was going on. If you knew who the company was, it'd be very obvious what was going on. Okay. But that what really happened is they kind of shifted from enterprise focus to HBF. And so that floats some things down in terms of even reviewing our tester. And now, then they came back to us in the summer and we're like, okay, here's the new tester we'd like. So, okay. Maybe that's good. It's, For people that, you know, you're tapping your fingers, it's taking a long time. But that's part of what happened there. But at this point, again, we, you know, we walked up. They're actually – they thought we were just going to take their wafer and stick it into one of, like, our NPs with a manual setup. And we showed them a fully integrated machine. So they walked up, and we put their wafer in a foop, put the foop onto the Sierra automated wafer pack aligner, ran the wafer, it opened up the blade, put the wafer in the wafer pack, put the wafer back in the blade, closed the blade, ran the chest, gave them the results.

Larry Shlebina | Analyst, Shlebina Capital:

It's pretty impressive. So you're ready to go for production. So it seems like they're going to need more capacity based on everything that's going on in the memory market.

Gane Erickson | President and Chief Executive Officer:

Exactly. And right now they're all flush with margins. How's that? Right? So I agree. You know what? You know, we've been – Larry, you, as people that follow Larry, is our greatest cheerleader, along with me, in memory strategy for us. We are spending money. It is part of, as Chris alludes to, you know, we could be doing better. Well, at these relevant levels, this is – you know, we're not happy with these relevant levels, right? We're not making money at these levels. But we would be making more money. we're spending money. We've got our foot on the gas, and in fact, it's our expectation that we'll increase the R&D spend, particularly in AI, wafer-level burden, a little bit in the package, because we spent a lot of money on that in just this last year for package, getting this new product out, and then the memory system, which will be a blade in our Fox system, basically.

Larry Shlebina | Analyst, Shlebina Capital:

It should be, it should pay off. That's So hopefully soon, sooner rather than later.

Gane Erickson | President and Chief Executive Officer:

I vote yes, too. As a shareholder, I think it's good money to be spent.

Larry Shlebina | Analyst, Shlebina Capital:

That's all I have. Thanks, Gary. Thank you, Larry.

Operator | Conference Operator:

Again, if there are any remaining questions, please indicate so by pressing star 1 on your touchtone phone. Okay. I'm showing no further questions in the queue. I would like to turn the call back to management for closing remarks.

Gane Erickson | President and Chief Executive Officer:

Thank you, operator. And thank you, everybody. We really appreciate you guys taking the time to spend an hour with us. I think about that exactly again. And we'll keep you guys updated. Stay tuned. We're really excited about this and hope that the orders will come in shortly enough to be able to make this less dramatic as we go forward and set us up for a really strong year heading into next year. So appreciate it. If you are in town, we are in Fremont, California, near Silicon Valley, give us a call, set something up, come by, take a look at the facility. If you haven't seen our tools, they're very impressive, and you can get a feel of the capacity because we have a lot of systems on the manufacturing line right now. So take care, and Happy New Year to everyone.

Operator | Conference Operator:

This concludes today's conference, and you may disconnect your lines at this time. Thank you for your participation.